

Implementation of pattern-mixture models in randomized clinical trials

By P. Bunouf⁽¹⁾ and G. Molenberghs⁽²⁾

(1) Laboratoires Pierre Fabre, France

(2) I-BioStat, Universiteit Hasselt and KU Leuven, Belgium

Summary. Modern analysis of incomplete longitudinal outcomes involves formulating assumptions about the missingness mechanisms and then using a statistical method that produces valid inferences under this assumption. In this manuscript, we define missingness strategies for analyzing randomized clinical trials (RCTs) based on plausible clinical scenarios. Penalties for drop-out are also introduced in an attempt to balance benefits against risks. Some missingness mechanisms are assumed to be non-future dependent (NFD) which is a subclass of missing-not-at-random. NFD stipulates that missingness depends on the past and the present information, but not on the future. Missingness strategies are implemented in the pattern-mixture modeling (PMM) framework using multiple imputation (MI) and it is shown how to estimate the marginal treatment effect. Next, we outline how MI can be used to investigate the impact of drop-out strategies in subgroups of interest. Finally, we provide the reader with some points to consider when implementing PMM-MI analyses in confirmatory RCTs. The data set that motivated our investigation comes from a placebo-controlled RCT design to assess the effect on pain of a new compound.

KEY WORDS : incomplete longitudinal outcome ; missing not at random ; non-future dependence ; pattern-mixture model ; multiple imputation.

1 Introduction

Clinical developers are becoming increasingly aware of good practices in analyzing incomplete longitudinal outcomes in randomized clinical trials (RCTs). Their analyses involve formulating assumptions about the missingness mechanisms and then using a statistical method that produces valid inferences under these assumptions. Consequently, the formulation of missingness assumptions in a transparent and interpretable manner has become a key aspect. A major cause of missingness in RCTs is drop-out. Missingness is at random (MAR) if drop-out occurrence is independent of missing outcome values, conditionally on the observed ones. If the covariates are fully observed, additional dependence on covariates is allowed for too. When MAR fails to hold, missingness is not at random (MNAR). MNAR implies that drop-out occurrence depends on an outside variable not in the model or is related to unobserved outcome values at the drop-out time and possibly afterwards, even when conditioned on available information. The consequence of MNAR is that missing outcome values cannot be reliably predicted using observed measurements (i.e., covariates and outcome values).

Pattern-mixture modeling (PMM) is a framework that can be considered when missingness is MNAR [1]. PMM stratifies the sample of subjects by missingness pattern and formulates distinct models to estimate parameters within each pattern. In RCTs with multiple scheduled visits, time is often modeled as a fixed class effect and patterns are defined based on drop-out at every visit. Whereas parameters are all identified in completers (i.e., subjects who complete the trial until the last visit), some parameters are unidentified in the patterns of drop-out subjects. This can be overcome by setting unidentified parameters equal to functions of the parameters describing the distributions of other patterns.

The so-called identifying restrictions indicate from which patterns information is borrowed. In Little's taxonomy [2], complete-case missing values (CCMV) stipulates that missing information is borrowed from completers. In neighboring-case missing values (NCMV), the closest neighboring pattern is used instead. NCMV implies that any unidentified parameter at a visit is estimated in the pattern of subjects having their last observed outcome value at this visit. Available-case missing values (ACMV) offers a compromise between CCMV and NCMV as all available patterns are used weighted by occurrence of each pattern. ACMV has a particular status since this is the natural counterpart of MAR in the PMM framework. In practice, analysis assuming MAR is often a point of departure for sensitivity analyses assuming MNAR.

Non-future missing values (NFMV) is another identifying restriction that offers appealing perspectives for sensitivity analysis since the user has full freedom to choose the distribution of the first unobserved (or present) outcome value given previous measurements. Under NFMV, missingness depends on the past and the present, but not on future unobserved outcome values. In other words, missingness is non-future dependent (NFD). NFD assumption is a sub-class of MNAR. The correspondence between NFMV and NFD allows the formulation of comprehensible drop-out strategies since drop-out mechanisms are directly characterized by the free distributions of present outcome values.

In this manuscript, we define several drop-out strategies based on plausible clinical scenarios. We suggest solutions to overcome major drawbacks of well known single-imputation concepts such as baseline-observation-carried-forward (BOCF) and last-observation-carried-forward (LOCF). In an attempt to balance benefits against risks in the assessment of the treatment effect, we introduce penalties which are intended to reflect the prejudice suffered by subjects due to drop-out. In the last set of drop-out strategies, we question the choice of the first unobserved visit to characterize the present in our NFD implementation.

Drop-out strategies are implemented in the pattern-mixture modeling (PMM) framework using multiple imputation (MI). We propose an approach to analyse the marginal treatment effect in the presence of moderate imbalance in the treatment allocations across patterns. MI-based methods offer a transparent way to represent missingness strategies [3]. In line with this, we investigate the impact of drop-out strategies in subgroups of interest using individual profiles obtained from the imputed values. Finally, we provide technical information to implement analysis using an existing freely-available SAS program [4].

To this end, we have re-analyzed a data set that comes from a placebo-controlled clinical trial to assess the effect on pain of a new compound [5][6]. A continuous visual analogue scale (VAS) was used to assess pain intensity and a binary outcome, whose values indicate clinical response or non-response, was derived. The next section describes the case study and outlines some points to consider during analysis preparation. Identifying restrictions and drop-out strategies are laid out in Section 3. In Section 4, we describe our approach to analyse the marginal treatment effect. Results obtained under drop-out strategies are contrasted in Section 5 whereas Section 6 and 7, respectively, contain the discussion and the concluding remarks. The appendix provides the information to implement analysis using software.

2 Case study

2.1 Trial design and statistical objectives

The data set was collected in a multi-country European phase-III double-blind clinical trial of which the objective was to compare the effect on pain intensity of a test product (Test) versus placebo (PCB) in subjects suffering from fibromyalgia. Four hundred thirty three (433) subjects were randomized to receive Test versus 447 subjects to receive PCB, that is 880 subjects in total.

After random assignment to either Test or PCB at the randomization visit (visit 0), subjects underwent a four-week escalation dose period followed by three four-week periods under stable dose. All periods ended at the scheduled monthly visits that are visit 1 at the end of dose escalation and visits 2–4 under stable dose. Subjects who discontinued were asked to undergo a specific visit at the discontinuation date, termed the drop-out visit.

Pain intensity was reported by subjects on a VAS that ranges from 0 (no pain) to 100 (worst pain imaginable) on a daily basis using an electronic diary. To obtain pain intensity levels by visit, these daily assessments were averaged over 2-week periods prior to the visits. The two-week period offers a reasonable compromise between one week which covers too short a period to be clinically meaningful and 4 weeks, which presents the disadvantage that, as pain intensity is recorded daily between a target visit and the preceding one, the assessments in the first two weeks are more related to the preceding visit than to the target visit. Last, the four values of a continuous longitudinal outcome were derived as the pain intensity level at randomisation minus the ones at visits 1–4. Thus, a positive value of outcome indicates a decrease in pain intensity.

The statistical properties of outcome were studied with respect to the averaging period length in [5]. It is especially mentioned that the mean and the variability of outcome values decrease as the averaging period length increases. The decrease of mean is caused by the global trend in the pain intensity to improve over time. So, any extension of the averaging period toward the preceding visit tends to involve worse pain assessments.

Next, the four values of a longitudinal binary outcome were derived in terms of clinical response based on a minimum 30% improvement of the pain level at randomisation on the continuous outcome values. In chronic pain diseases, including fibromyalgia, this gain is regarded as clinically relevant. For the drop-out subjects, the pain intensity level was assessed at the drop-out visit, and the corresponding values of the continuous and the binary outcomes were repositioned to the next scheduled visit for analysis.

The estimation of treatment effect at visit 4 was the primary objective of the trial. The original analyses, as mentioned in the trial protocol, were based on BOCF and LOCF principles to impute missing outcome values. The continuous outcome values were fitted using a covariance analysis model and the binary outcome values were fitted using a logistic regression model. Pain level at randomization was a pre-specified key covariate whereas country was a pre-specified stratum factor. The objective of the analyses described in this manuscript is to assess the treatment effects at visit 4 adjusted on the same covariate and stratum factor under a set of drop-out strategies. Analyses are conducted on all randomized subjects.

2.2 Causes of missingness

Out of the 880 randomized subjects, 77/447 subjects (17.2%) in the PCB group and 126/433 subjects (29.1%) in the Test group dropped out from the trial. In our case study, drop-out from the trial also means drop-out from the follow-up even if treatment intake had stopped earlier. Three subjects who dropped out very early have no assessment after randomization. All the other drop-out subjects have completed their drop-out visit and reported their pain intensity level. There are only two ‘intermittent’ missing pain levels (i.e., pain levels missing not due to drop-out). The first occurred at the randomization visit because of a technical problem with the electronic diary whereas the second occurred at visit 3 and was caused by the absence of a subject from home.

The distribution of subjects by drop-out cause by treatment group is given in Table 1.

TABLE 1 – Frequencies (percentages) of subjects by drop-out cause by treatment group

Causes	PCB	Test
Adverse event (AE)	44 (9.8)	96 (22.2)
<i>Serious AE due to treat.</i>	1 (0.2)	3 (0.7)
<i>AE due to treatment</i>	30 (6.7)	89 (20.6)
<i>Serious AE not due to treat.</i>	3 (0.7)	0 (—)
<i>AE not due to treatment</i>	10 (2.2)	4 (0.9)
Other than AE	33 (7.4)	30 (7.0)
<i>Patients’s decision</i>	11 (2.5)	11 (2.5)
<i>Investigator’s decision</i>	0	4 (0.9)
<i>Therapeutic failure</i>	18 (4.0)	10 (2.3)
<i>Other reasons</i>	4 (0.9)	5 (1.2)
Total	77 (17.2)	126 (29.1)

The main cause of drop-out was adverse event (AE). The higher drop-out rate for Test stems from a higher occurrence of AEs in this group. Particularly, there are three times more subjects who dropped out for AE due to the treatment in the Test group than in the PCB group (6.7% for PCB versus 20.6% for Test). Among the 119 concerned AEs, most of them were pre-identified and particularly under investigation. These can be gathered into four classes :

- ▷ Cardio-vascular disorders including tachycardia, hypertension, and heart rate increase with 2/447 (0.4%) for PCB versus 25/430 (5.8%) for Test,
- ▷ Gastrointestinal disorders including nausea and vomiting with 4/447 (0.9%) for PCB versus 12/430 (2.8%) for Test,
- ▷ Skin reactions including rash and hyperhydrosis with 1/447 (0.2%) for PCB versus 15/430 (3.5%) for Test,
- ▷ Nervous system disorders including dizziness and headache with 6/447 (1.3%) for PCB versus 15/430 (3.5%) for Test.

Conversely, subjects who dropped out for other causes than AE are similarly distributed in the two groups with 7.4% for PCB and 7.0% for Test. In this category, therapeutic failure concerns twice less subjects under Test (2.3%) than under PCB (4.0%). Further investigations (not shown in Table 1) reveal that the earlier the drop-out, the greater the rate of drop-out caused by AE (from 77.8% if drop-out occurs between randomization and visit 1 to 40.6% between visit 3 and visit 4). The distribution of drop-out causes in our case study is classical of many clinical trials in chronic pain.

2.3 Characteristics of patterns

Patterns are defined based on drop-out at every visit. Accordingly, patterns 0–4 respectively consist of subjects with 0 to 4 available outcome values. Pattern 0 contains the subjects without pain level at randomisation and the three subjects without pain level after randomization. Pattern 4 contains the subjects with complete outcome profile. However, this pattern does not allow distinguishing between the completers and the subjects who dropped out after visit 3 and have their drop-out visit repositioned to visit 4. To clarify this, we define a pattern 5 which contains the completers, whereas pattern 4 keeps the subjects who dropped out after visit 3. This separation does not contradict the pattern definition given hereabove since one may consider that the completers actually dropped out at a virtual visit 5 after the end of the trial follow-up.

Figure 1 displays the unadjusted mean profiles per pattern by treatment group.

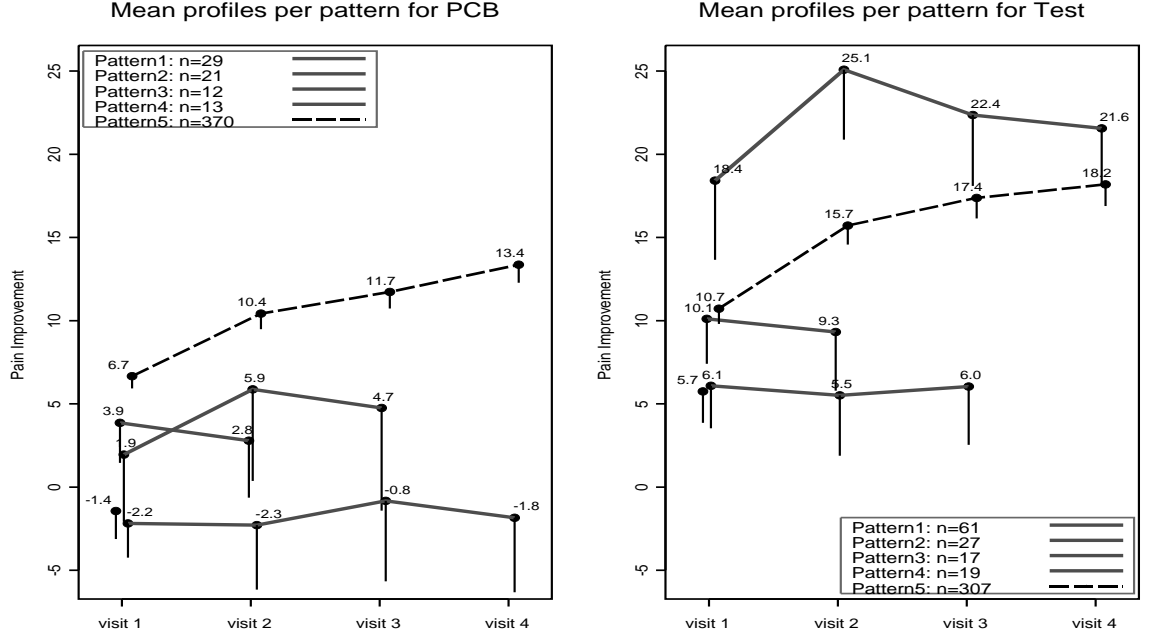


FIGURE 1 – Unadjusted mean profiles and frequencies per pattern by treatment group.

We first observe a substantial difference between treatment groups in the means per pattern at visit 1. This is caused by the early response to the treatment during the escalation dose period (from randomization to visit 1) which is larger under Test than under PCB. In subjects who stay longer on the trial, the mean profiles of drop-out patterns exhibit a stagnation until the drop-out visit.

Conversely, the pattern of completers (pattern 5) shows sustained improvement until visit 4 in the two treatment groups. It is worthwhile noting that the mean profiles in patterns 4 exhibit a huge difference between treatment groups. This difference is partly explained by the causes of drop-out, which are less related to AE and probably more to efficacy, in this drop-out pattern than in the others. The difference in the mean profiles between pattern 4 and pattern 5 supports the decision of separating them.

2.4 Pattern sizes by stratum

Pattern sizes by stratum is an important point to consider before implementing a stratified PMM analysis. The reason is that parameter estimation can fail if a minimum number of values are not available. In our case study, pattern 0 is not used for parameter estimation but is pooled with pattern 1 for imputation. The imputation models are fully described in

Section 4.1. Particularly, the stratum factor country and the full group-by-visit interaction are specified as fixed factors. In this setting, parameter estimation requires a minimum of three available outcome values per pattern by country by treatment group. As the trial was conducted in twelve European countries, this pre-requisite is not reached and a tricky pooling of countries must be done.

The trial was conducted in the Czech republic (CZ ; $n=55$), Denmark (DK ; $n=17$), Finland (FI ; $n=60$), France (FR ; $n=184$), Germany (GE ; $n=42$), Italy (IT ; $n=101$), Norway (NO ; $n=53$), Poland (PL ; $n=31$), Portugal (PT ; $n=16$), Romania (RO ; $n=59$), Spain (ES ; $n=111$), and Sweden (SE ; $n=151$). A first pooling of countries into four regions was done based on their geographical proximity. These regions are central Europe (CZ, GE, PL, RO), southern Europe (ES, IT, PT), France (FR), and the Nordic countries (DK, FI, NO, SE). Then, central and southern Europe were pooled into the same region to reach the minimum pattern size. The consequence of this pooling is that the stratum factor country is replaced by the three-class factor ‘region’ in the imputation models.

The pattern sizes by region by treatment group are provided in Table 2.

TABLE 2 – Pattern sizes by region by treatment group

		central + south	France	nordic countries	All Countries
Patterns 0+1	PCB	1+11	9	1+9	2+29
	Test	1+28	10	1+23	2+61
Pattern 2	PCB	9	4	8	21
	Test	11	5	11	27
Pattern 3	PCB	4	4	4	12
	Test	3	7	7	17
Pattern 4	PCB	4	3	6	13
	Test	11	3	5	19
Pattern 5	PCB	187	70	113	370
	Test	145	69	93	307
All Patterns	PCB	216	90	141	447
	Test	199	94	140	433
	All Groups	415	184	281	880

3 Identifying restrictions and drop-out strategies

3.1 Description of identifying restrictions

The basis of pattern-mixture modeling results from a particular decomposition of the joint distribution of the outcome variable together with the drop-out indicator. The pattern-mixture distribution of the complete outcome values y_1, \dots, y_T is given by :

$$f(y_1, \dots, y_T) = \sum_{t=1}^T \alpha_t f_t(y_1, \dots, y_T), \quad (1)$$

where α_t denotes the probability of pattern t and $f_t(y_1, \dots, y_T)$ stands for $f(y_1, \dots, y_T|t)$.

In our case study, patterns are defined based on drop-out at every visit. More precisely, if the t th outcome value is the last observed one in a subject, this subject belongs to pattern t , $t = 1, \dots, T$. In (1), the distribution of the whole population is expressed in terms of a mixture of the distributions of pattern populations. These, in turn can be decomposed as :

$$\begin{aligned} f_t(y_1, \dots, y_T) &= f_t(y_1, \dots, y_t) f_t(y_{t+1}, \dots, y_T | y_1, \dots, y_t) \\ &= f_t(y_1, \dots, y_t) \prod_{s=t+1}^T f_t(y_s | y_1, \dots, y_{s-1}). \end{aligned} \quad (2)$$

The first component in (2) is identified from the observed outcome values. The second is a product of conditional pattern distributions, which are unidentified since the values of y_s are unobserved in these patterns. This can be overcome by setting unidentified parameters equal to functions of the parameters describing the distributions of other patterns. The identifying restrictions, informally introduced in Section 1, are used to this effect.

Under CCMV, identification is based on pattern T , the pattern of completers. This can be formalized by :

$$f_t(y_s | y_1, \dots, y_{s-1}) = f_T(y_s | y_1, \dots, y_{s-1}), \quad s = t + 1, \dots, T. \quad (3)$$

Under NCMV, the neighboring pattern is used instead :

$$f_t(y_s | y_1, \dots, y_{s-1}) = f_s(y_s | y_1, \dots, y_{s-1}), \quad s = t + 1, \dots, T. \quad (4)$$

Identification can also be based on all identified patterns as specified in the formulation :

$$f_t(y_s | y_1, \dots, y_{s-1}) = \sum_{j=s}^T \omega_{sj} f_j(y_s | y_1, \dots, y_{s-1}), \quad s = t + 1, \dots, T. \quad (5)$$

Every convex set of ω_{sj} 's that sums to 1 provides a valid identification scheme. In [8], the ω_{sj} 's are determined such that (5) corresponds with ACMV.

Alternatively, NFMV offers the appealing characteristic that the distributions of present values, that we denote by g_t for the sake of clarity, are left unconstrained [9]. In our analysis, we set the g_t 's equal to their f_{t+1} counterparts, in the spirit of NCMV, with a possibility to shift the mean by a value Δ . Formally, this can be expressed by :

$$g_t(y_{t+1}|y_1, \dots, y_t) = f_{t+1}(y_{t+1} + \Delta|y_1, \dots, y_t). \quad (6)$$

Future values are then imputed, under ACMV, conditionally on the past and the present.

In our case study, the closest neighboring patterns involved in NCMV and NFMV are all drop-out patterns. Indeed, pattern 5 is never used for identification since pattern 4 does not contain incomplete outcome profiles. Of note, the alternative choice of basing NFMV identification on the pattern of completers, in the spirit of CCMV, would have been inappropriately optimistic. As described in Figure 1, the mean profiles of completers show regular improvements in the two treatment groups, whereas the mean profiles of the drop-out patterns exhibit stagnation over time. Moreover, the large size of patterns 5 would impose an inappropriately low uncertainty for parameter estimation.

The correspondence between NFMV and NFD allows formally relating drop-out occurrence to the g_t 's. Through (6), we state that drop-out is caused by, or associated with, a mean decrease by Δ of the present values relative to the values observed in subjects who dropped out at this visit, without other possible causes, or associations, involving future values.

3.2 Description of drop-out strategies

Although MAR is impossible to demonstrate [10], primary analysis in RCTs often relies on this assumption. As laid out in [7], MAR is compatible with *de jure* (or per-protocol) analysis of which an objective is to estimate the treatment effect as if a drop-out subject continued the trial under the same conditions until the last scheduled visit. A consequence in our case study is that the imputation parameters are estimated using all subjects in the two treatment groups. This rule is applied under ACMV, which is the counterpart of MAR, as well as under CCMV and NCMV, which are often used to study the effect of a certain departure from MAR. These three identifying restrictions characterize the *de jure* analyses.

Alternatively, a *de facto* (or intention-to-treat) analysis aims to reflect the effect of the treatment as well as the impact of treatment withdrawal subsequent to drop-out. In our case study, we assume that subjects under PCB continue on PCB after drop-out and subjects

under Test switch to PCB. Consequently, the imputation parameters are estimated using subjects of the PCB group only. This approach is referred to as *jump to reference* (J2R) in earlier work [17].

De facto analyses are characterized by drop-out strategies based on NFD. We first introduce N-LOCF which is intended to overcome a major drawback of LOCF single imputation by allowing an appropriate degree of uncertainty. Under N-LOCF, the present values are imputed without shifting subjects' distribution mean (i.e., $\Delta = 0$). So, we expect that a drop-out subject behaves as a subject of the neighboring pattern in the PCB group with the same previous measurements. Future values are then imputed conditionally on the past and the present, assuming PCB-MAR. In the same vein, we define N-BOCF which takes up the concept of BOCF single imputation. Under N-BOCF, we set the values of Δ equal to minus the individual gains observed from randomization to the drop-out visit.

The following drop-out strategies rely on clinical rationales. In an attempt to balance benefits against risks in the assessment of the treatment effect, we now introduce penalties which are supposed to reflect the prejudice suffered by subjects due to drop-out. Under N-DO5, any drop-out is associated with a penalty set to $\Delta = -5$. This value was discussed and approved for its clinical meaningfulness by experts in chronic pain. This also corresponds to minus the expected treatment effect under the alternative in the trial protocol.

The drop-out strategy N-AE5 refines N-DO5 by incorporating information about the cause of drop-out. Under N-AE5, missing outcome values are imputed, assuming :

1. NFD($\Delta = -5$) if a subject drops out for AE due to the treatment,
2. NFD($\Delta = -10$) if a subject drops out for serious AE due to the treatment,
3. N-LOCF or NFD($\Delta = 0$) if a subject drops out for AE not due to the treatment,
4. MAR if a subject drops out for other reasons than AE.

Since the drop-out rate is greater in the Test group, the greater the penalty, the greater the conservatism introduced by NFD into the analysis. The next two drop-out strategies aim at measuring the impact of penalty increase. If subjects dropped out for AE due to the treatment, the penalty value is brought to $\Delta = -10$ in N-AE10 and $\Delta = -15$ in N-AE15. The other rules in N-AE5, i.e., the doubling of the Δ value if AE is serious and the zeroing if AE is not due to the treatment, are kept in N-AE10 and N-AE15.

In the last set of drop-out strategies, we question the choice of the first unobserved visit to characterize the present in our NFD implementation. Indeed, the drop-out visits are repositioned to the next scheduled visit for analysis. So, in many drop-out subjects the first

unobserved visits may be likened to a (near) future. Moreover, the influence of the cause of drop-out, such as an AE, on a subject's assessment is not well characterized. Thereby, we cannot rule out that the outcome value at a drop-out visit combines some aspects of efficacy and safety. In an attempt to address this, we introduce N-AE5-L, N-AE10-L, and N-AE15-L, which copy the penalty rules applied in N-AE5, N-AE10, and N-AE15, respectively. However, the drop-out visits are removed from analysis in the subjects who dropped out for AE.

In Section 5, we compare the results obtained from the de jure and de facto analyses to those obtained after BOCF and LOCF single imputations, as well as in completers and in subjects with complete outcome profiles (i.e., in subjects who dropped out after visit 3). All analyses are conducted using the same statistical models, which are described in the next section. For BOCF and LOCF analyses specifically, the three subjects of pattern 0 without pain level after randomization are imputed with a 0 value for the continuous outcome and a clinical non-response. The fourth subject of pattern 0 without pain level at randomization is excluded from analyses.

Table 3 sums up the drop-out strategies described in this section.

TABLE 3 – Description of drop-out strategies

Type of analysis	Drop-out strategy	Missingness mechanisms	Type of imputation	Parameter estimation	Data set
—	BOCF	Unknown	Single	—	All visits
—	LOCF	Unknown	Single	—	All visits
—	Complete	—	—	—	Complete profiles
—	Completers	—	—	—	Completers
de jure	NCMV	Unknown	Multiple	All subjects	All visits
de jure	ACMV	MAR	Multiple	All subjects	All visits
de jure	CCMV	Unknown	Multiple	All subjects	All visits
de facto	N-BOCF	NFD	Multiple	PCB group	All visits
de facto	N-LOCF	NFD	Multiple	PCB group	All visits
de facto	N-DO5	NFD	Multiple	PCB group	All visits
de facto	N-AE5	NFD/MAR	Multiple	PCB group	All visits
de facto	N-AE5-L	NFD/MAR	Multiple	PCB group	Scheduled visits
de facto	N-AE10	NFD/MAR	Multiple	PCB group	All visits
de facto	N-AE10-L	NFD/MAR	Multiple	PCB group	Scheduled visits
de facto	N-AE15	NFD/MAR	Multiple	PCB group	All visits
de facto	N-AE15-L	NFD/MAR	Multiple	PCB group	Scheduled visits

4 Implementation of PMM analysis using MI

4.1 General features

The PMM analyses are implemented following the standard MI approach, as described in [11], which includes pattern parameters estimation, missing values imputation, and pooled analysis. Some other basic features follow general recommendations for MI analysis, which are stated, justified, and exemplified in [6]. These are :

1. MI on the original scale followed by analysis of the desired derived outcome is a more informed strategy than direct analysis of the derived outcome ;
2. Analysis at the desired timepoint provides valid inferences if all the effects are properly fitted over time by the imputation models ;
3. MI offers a transparent way to represent the impact of drop-out strategies.

In our case study, missing values are imputed on the continuous scale in accordance with recommendation 1. For the binary outcome, we reuse the data sets imputed on the continuous scale to derive the clinical responses or non-responses by subject by imputation.

The imputation models used are mixed models for repeated measures (MMRM) with the three-class factor region and the full baseline-by-visit and group-by-visit interactions for the fixed effects and unstructured error covariance matrix. The baseline-by-visit interaction considers possible variation of the impact of baseline over visits. Provided the study is large enough, such adjustment allows minimum assumption in the imputation stage, so as not to compromise the analysis stage. To analyze the outcome values at visit 4, we use a covariance analysis model for the continuous outcome and a logistic regression model for the binary outcome as originally planned in the trial protocol. Both analysis models incorporate the pain level at randomization as covariate and the three-class factor region as stratum factor. According to recommendation 2, these models to analyze outcome values at visit 4 provide as valid inferences as longitudinal analysis models congenial to the imputation models.

We implement recommendation 3 using summary outcome values by subject which are directly obtained from the imputed values. The summary continuous values are simply the means by subject over imputations. For the summary binary values, the clinical responses or non-responses are derived by subject by imputation. The subject is declared as clinical responder if at least half of values correspond to clinical responses. The summary values are used to investigate, via the analysis models described hereabove, the impact of the drop-out strategies in the subgroups of subjects who dropped out for AE and those who

dropped out for other reasons. The summary binary values are also used for illustrative purpose to provide raw frequencies of clinical responders.

Unlike pooled analysis in Rubin’s method, an analysis based on summary outcome ignores between-imputation variance. The relative increase of variance due to missing data depends on the ratio of the between-imputation variance to the total variance [11]. The maximum value for this criterion observed across drop-out strategies in our case study was 0.17. This indicates that the magnitude of ignored variance information is moderate. This also supports the conclusions of subgroup analyses using the summary outcomes.

4.2 Treatment-effect inferences

In this section, we describe a procedure in three stages to estimate the marginal treatment effect in a PMM-MI analysis. It is important to recall that the original Rubin’s method is conditional on patterns by construction. To obtain the marginal effect, the pattern-specific effects must be combined into a pattern-average effect. Some aspects of the procedure are further detailed in [4] and [12]. In the Appendix, we provide information to implement analysis using a SAS program which combine R functionalities.

4.2.1 Pattern parameter estimation

Distinct models are formulated within each pattern. Let us denote by $\mathbf{Y}_i = (y_{i,1}, \dots, y_{i,T})$ the complete outcome vector in the i th subject of pattern t and $\mathbf{Y}_{i,obs} = (y_{i,1}, \dots, y_{i,t})$ its observed part. The MMRMs per pattern can be expressed as :

$$\mathbf{Y}_{i,obs} = X_i \boldsymbol{\beta}_t + \boldsymbol{\epsilon}_i, \quad (7)$$

where $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \Sigma_t)$, Σ_t is unstructured, and the $\boldsymbol{\epsilon}_i$ ’s are independent. The matrix X_i contains the known subject covariates whereas $\boldsymbol{\beta}_t$ contains the unknown fixed effects. When the estimation of parameters is based on the PCB group only, the fixed factor treatment, and related interactions, are omitted. This first stage is aimed at estimating the pattern parameters $\boldsymbol{\beta}_t$ and Σ_t . Estimators are denoted by $\hat{\boldsymbol{\beta}}_t$ and $\hat{\Sigma}_t$ in what follows.

4.2.2 Imputation

The process of imputation is conducted sequentially by value. The identifying restriction indicates from which patterns missing information is borrowed. Whereas imputations under

CCMV (3) and NCMV (4) use unique patterns, imputations under ACMV (5) and NFMV (6) are based on several patterns.

We now describe how to obtain a run of M imputed values of $y_{i,t+1}$ given $y_{i,1}, \dots, y_{i,t}$. Multiple imputation of $y_{i,t+2}, \dots, y_{i,T}$ follows the same process by considering the previous imputed values as observed ones. In our illustration, we suppose that $y_{i,t+1}$ is imputed from pattern r ($t+1 \leq r \leq T$). Let us denote by $\boldsymbol{\mu}_{i,r}$ the mean of \mathbf{Y}_i , which is $\boldsymbol{\mu}_{i,r} = X_i \boldsymbol{\beta}_r$. Based on appropriate parts of $\boldsymbol{\mu}_{i,r}$ and Σ_r , we further define the distributions of $\mathbf{Y}_{i,obs} \sim N(\boldsymbol{\mu}_{i,r,1}, \Sigma_{r,11})$ and $y_{i,t+1} \sim N(\mu_{i,r,2}, \Sigma_{r,22})$. Their covariance terms are denoted by $\Sigma_{r,12}$ and $\Sigma_{r,21}$. Using $2|1$ as notation for $y_{i,t+1}|y_{i,1}, \dots, y_{i,t}$, the conditional pattern distribution of $y_{i,t+1}$ given $y_{i,1}, \dots, y_{i,t}$ is described by :

$$f_r(y_{i,t+1}|y_{i,1}, \dots, y_{i,t}) \sim N(\mu_{i,r,2|1}, \Sigma_{r,2|1}),$$

where

$$\begin{aligned} \mu_{i,r,2|1} &= \mu_{i,r,2} + \Sigma_{r,21}[\Sigma_{r,11}]^{-1}(\mathbf{Y}_{i,obs} - \boldsymbol{\mu}_{i,r,1}), \\ \Sigma_{r,2|1} &= \Sigma_{r,22} - \Sigma_{r,21}[\Sigma_{r,11}]^{-1}\Sigma_{r,12}. \end{aligned} \tag{8}$$

Uncertainty pertaining to the parameters $\boldsymbol{\beta}_r$ and Σ_r is now incorporated through their Bayesian distributions. On the basis of the non-informative Jeffreys' priors in the Gaussian context, the values of $\widehat{\boldsymbol{\beta}}_r^{(m)}$ and $\widehat{\Sigma}_r^{(m)}$, $m = 1, \dots, M$, are randomly drawn from their respective posterior predictive distributions. After the derivation of $\widehat{\boldsymbol{\mu}}_{i,r}^{(m)}$, the imputed values of $y_{i,t+1}$ are drawn from the conditional pattern distributions :

$$f_r^{(m)}(y_{i,t+1}|y_{i,1}, \dots, y_{i,t}) \sim N(\widehat{\boldsymbol{\mu}}_{i,r,2|1}^{(m)}, \widehat{\Sigma}_{r,2|1}^{(m)}), \quad m = 1, \dots, M. \tag{9}$$

This stage requires the outcome value at visit 1 to be available. In our case study, this value is missing in the four subjects of pattern 0, and must be multiply imputed to begin with. The outcome value at visit 1 is derived from the pain levels at randomization (covariate) and visit 1. In one subject with missing covariate, the covariate is imputed from its conditional distribution given the pain level at visit 1 in all subjects. In the three other subjects of pattern 0 who have missing pain level at visit 1, the missing outcome value at visit 1 is imputed from its conditional distribution given the covariate in pattern 1 under NCMV, in all subjects under ACMV, and in pattern 5 under CCMV. Next, the only intermittent missing outcome value after randomization, which occurs in a subject of pattern 5 at visit 3, is multiply imputed from its conditional distribution given the observed outcome values in pattern 5.

4.2.3 Pooled analysis

Direct pooled analysis combines the treatment-effect estimates $\hat{d}^{(m)}$ obtained in imputations $m, m = 1, \dots, M$. According to Rubin's rule [11], the pooled estimate \bar{d} is the mean of $\hat{d}^{(m)}$ and the pooled variance is given by :

$$V = \overline{W} + \left(\frac{M+1}{M} \right) B, \quad (10)$$

where \overline{W} is the mean of $d^{(m)}$'s variances and B is the between-imputation variance such that :

$$B = \frac{1}{M-1} \sum_{m=1}^M (\hat{d}^{(m)} - \bar{d})^2.$$

To test the hypothesis of no treatment effect, the statistic \bar{d}/\sqrt{V} approximately follows a t -Student distribution with degrees-of-freedom equal to $r = (M-1)[1 + W/(1 + M^{-1})B]^2$.

Direct pooled analysis provides consistent treatment-effect estimate. However, inferences are conditional on pattern probabilities by construction. This conditioning is inappropriate in our context of a confirmatory clinical trial wherein inferences on marginal effects are expected.

For this purpose, we average the treatment effects across patterns with respect to pattern probabilities. Let us introduce the MT -dimensional vector :

$$\mathbf{d} = (\hat{d}_1^{(1)}, \dots, \hat{d}_1^{(M)}, \dots, \hat{d}_T^{(1)}, \dots, \hat{d}_T^{(M)}),$$

where $\hat{d}_t^{(m)}$ is the estimate in pattern t and imputation m . Using Rubin's rule, we first derive the vector of pooled estimates per pattern which is denoted by $\mathbf{d}^* = (d_1^*, \dots, d_T^*)$ in what follows. So, the d_t^* 's are the means of the $\hat{d}_t^{(m)}$'s per pattern and V^* is the $T \times T$ -dimensional total covariance matrix of \mathbf{d}^* which is given by :

$$V^* = W^* + \left(\frac{M+1}{M} \right) B^*,$$

where W^* is the diagonal matrix whose coefficients are the means of the d_t^* 's variances and B^* is the correlation matrix of \mathbf{d}^* .

Then, the pattern-specific information is combined with respect to the multinomial distribution of patterns. To this end, we use Rubin's rule a second time to derive the pattern-average parameters which are denoted by d^\dagger , W^\dagger , and B^\dagger , respectively, in what follows. Let us define the vector $\boldsymbol{\Pi}$ that contains the pattern probabilities, the marginal treatment effect is given by :

$$\mathbf{d}^\dagger = \boldsymbol{\Pi}' \mathbf{d}^*. \quad (11)$$

The total variance takes the form :

$$V^\dagger = W^\dagger + \left(\frac{T+1}{T} \right) B^\dagger, \quad (12)$$

where $W^\dagger = \mathbf{\Pi}' W^* \mathbf{\Pi} + \mathbf{d}^{*'} \text{Var}(\mathbf{\Pi}) \mathbf{d}^*$ and $B^\dagger = \mathbf{\Pi}' B^* \mathbf{\Pi}$. These formulas were derived and justified in [12]. To test the hypothesis of no treatment effect, the statistic $d^\dagger / \sqrt{V^\dagger}$ approximately follows a t -Student distribution with a degrees-of-freedom equal to $(T-1) * \{1 + W^\dagger / [(1 + T^{-1}) * B^\dagger]^2\}$.

As mentioned in [13], it is important to recall that averaging regression coefficients (such as treatment effect) over pattern assumes that covariates (such as treatment assignment) are equally distributed across patterns, otherwise estimators are biased. In other words, (11) assumes homogenous pattern probabilities in the two treatment groups. This assumption is questionable in our case study. Figure 1 indicates that the pattern probabilities are $\hat{\mathbf{\Pi}}'_P = (.065, .047, .027, .029, .831)$ for PCB and $\hat{\mathbf{\Pi}}'_T = (.141, .063, .039, .044, .712)$ for Test and the classical χ^2 to test homogeneity between treatment groups yields $p < 0.01$.

To tackle this issue, we introduce the idea of replacing d^\dagger in the test of the marginal effect by \bar{d} the point estimate of the pattern-conditional effect. Although this approach may appear somewhat inelegant as the statistic $\bar{d} / \sqrt{V^\dagger}$ is based on two terms given by different methods, it offers a pragmatic solution in presence of moderate imbalance in pattern probabilities between treatment groups. The justification rests upon the two main goals of analysis. The first is to provide unbiased point estimate. However, \bar{d} is consistent. The second is to draw inferences without type 1 error inflation. Under homogeneity assumption, the test of the marginal effect is more conservative than the test of the pattern-conditional effect as the integration of the distribution of patterns increases the variance V^\dagger in (12) relative to V in (10). Thus, our approach merely consists of using the appropriate terms to fulfill the two conditions given above. To unambiguously identify it in what follows, this approach is named *corrected marginal approach*.

All the results shown in the next section are obtained from the corrected marginal approach. To allow comprehensive interpretation, we also provide the results based on the full analyses of the pattern-conditional and the marginal effects in Section 5.3. This analysis scheme can be seen as a small, perhaps informal, attempt to formalize a set of primary and sensitivity analyses.

5 Results

5.1 Continuous outcome analysis

Figure 2 displays treatment-effect estimates obtained with the continuous outcome at visit 4 under the drop-out strategies specified in Table 3. We provide the differences between treatment groups with their 95% confidence limits and p -values. Results by treatment group are expressed in terms of adjusted means with standard errors.

As expected, owing to the greater drop-out rate in the Test group, the treatment effect estimated after LOCF imputation [4.30 (1.41) $p=0.002$] lies between that obtained after BOCF imputation [2.85 (1.37) $p=0.038$] and in subjects with complete outcome [5.87 (1.61) $p<.0001$]. Relative to BOCF, LOCF increases the means by treatment group by 0.1 for PCB and 1.6 for Test, suggesting that drop-out subjects under Test benefit from the treatment. Next, the treatment effect estimated in completers [5.14 (1.66) $p=0.002$] is lower than that obtained in subjects with complete outcome. This result was expected since subjects with complete outcome consist of the subjects of pattern 4 and the completers, and Figure 1 exhibits a huge difference between treatment groups in pattern 4.

Among PMM analyses, the results of de jure analyses exhibit weak variations. As expected, CCMV which uses the pattern of completers for imputation produces larger means by treatment group (11.9 for PCB and 16.9 for Test) than ACMV which uses all patterns (11.7 for PCB and 16.4 for Test) and NCMV which uses the closest neighboring patterns (10.7 for PCB and 15.6 for Test). CCMV also produces a larger treatment effect [5.03 (1.63) $p=0.002$] than ACMV [4.78 (1.63) $p=0.004$] and NCMV [4.85 (1.67) $p=0.004$].

The de facto analyses exhibit greater divergences. The treatment effect estimated under N-LOCF [3.92 (1.72) $p=0.024$] is related to that obtained with the single-imputation counterpart LOCF whereas N-BOCF [1.52 (1.89) $p=0.424$] provides a lower estimate than that obtained with BOCF. The allowance for appropriate uncertainty in the NFD strategies increases standard errors and p -values. Unlike BOCF, the treatment effect under N-BOCF is not statistically significant.

Owing to the greater drop-out rate in the Test group, the penalty for drop-out $\Delta = -5$ decreases by 0.71 the treatment effect estimated under N-DO5 [3.21 (1.78) $p=0.073$] relative to N-LOCF which involves $\Delta = 0$. The inferences produced by N-AE5 [3.13 (1.74) $p=0.0073$] are very close to those of N-DO5 although the penalty is zeroed if subjects dropped out for other reasons than AE. Indeed, these subjects are similarly distributed in the two groups (7.4% for PCB and 7.0% for Test based on Table 1) so that the means by

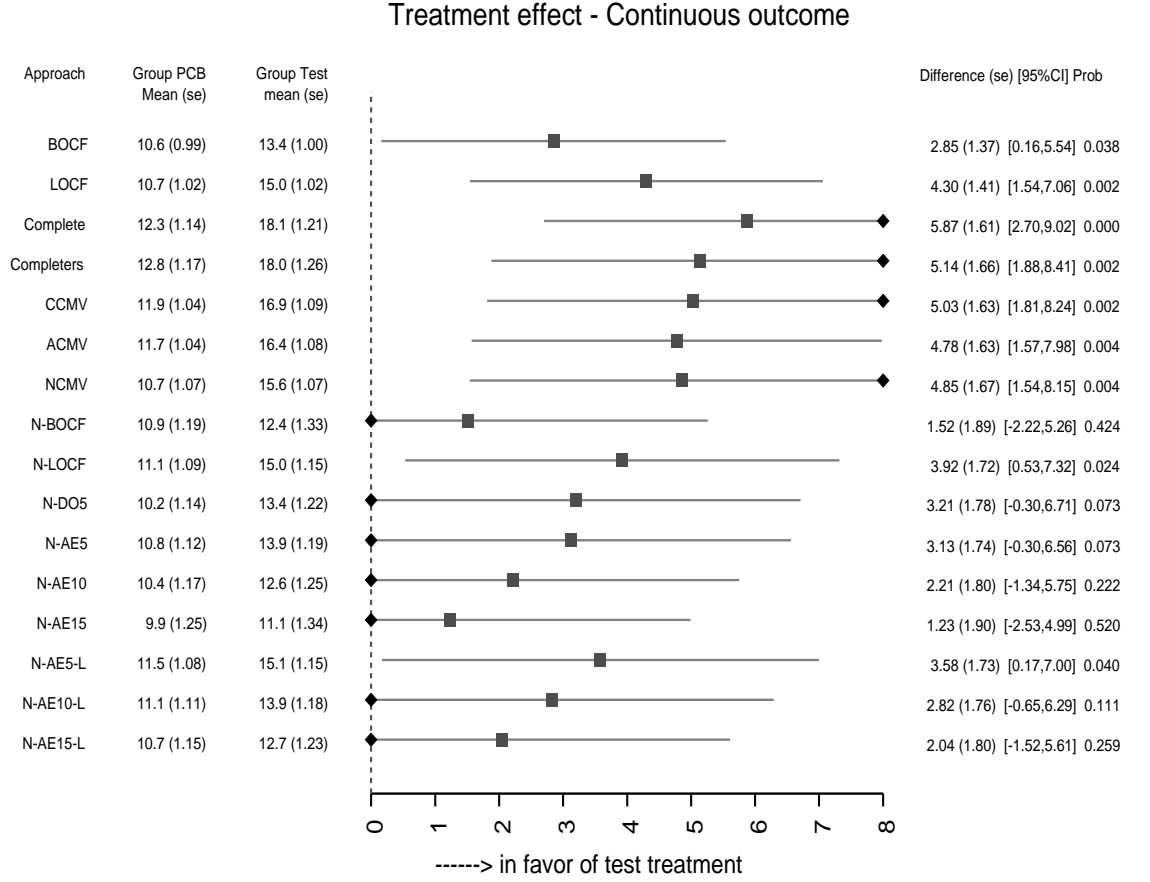


FIGURE 2 – Treatment-effect estimates obtained with the continuous outcome at visit 4.

treatment group estimated under N-AE5 increase similarly (+0.6 for PCB and +0.5 for Test). As a global rule for the N-AEx strategies, the penalty increase makes the treatment-effect estimate decreasing, as observed under N-AE10 [2.21 (1.80) $p=0.222$] and N-AE15 [1.23 (1.90) $p=0.520$].

For the N-AEx-L strategies, the removal of the last visit if drop-out is caused by an AE increases the means by treatment group relative to the N-AEx counterparts. This impact was expected since Figure 1 exhibits mean profiles in the drop-out patterns which slightly decline at the last visit. The removal of the last visit also increases the separation between treatment groups under N-AE5-L [3.58 (1.73) $p=0.040$], N-AE10-L [2.82 (1.76) $p=0.111$], and N-AE15-L [2.04 (1.80) $p=0.259$].

Several authors have suggested the search of a penalty value applicable to drop-out subjects at which conclusions change from being favorable for Test to being unfavorable [14]. In our case study, we have conducted such investigation under the strategies N-AEx and N-AEx-L. The limit values which zero the treatment effect (i.e., no difference between treatment groups) are $\Delta = -21$ under N-AEx and $\Delta = -30$ under N-AEx-L. In the same

vein, the limit values which cancel the statistical significance (i.e., $p=0.05$) are $\Delta = -3$ under N-AEx and $\Delta = -6$ under N-AEx-L. These limit values can be interpreted as the maximum price to pay for drop-out so that Test remains beneficial to subjects in the trial. If the penalty value is below these limit values, the risks for subjects do not balance the benefits of Test. Of note, several authors have also suggested to introduce conservatism, such as in the δ -method [15], by deliberately penalizing subjects of the Test group. In our analysis, the penalties are applied to drop-out subjects whatever their treatment groups. Ultimately, this allows a sensible representation of true treatment effects, underpinning a main trial objective [16].

5.2 Binary outcome analysis

Figure 3 displays treatment-effect estimates obtained with the binary outcome at visit 4 under the drop-out strategies specified in Table 3. We provide odds-ratios with their 95% confidence limits and p -values. Results by treatment group are described with frequencies and percentages of clinical responders based on the summary binary values.

The inferences obtained after BOCF [1.37 (0.15) $p=0.029$] and LOCF [1.53 (0.14) $p=0.003$] imputations confirm the major trends observed with the continuous outcome. The greater odds-ratio value obtained after LOCF imputation relative to BOCF is caused by 16 additional clinical responders for Test against 5 for PCB. Despite the change of outcome scale, the p -values obtained after BOCF and LOCF imputations, as well as in subjects with complete outcome [1.74 (0.16) $p<0.001$] and in completers [1.63 (0.16) $p=0.002$], are surprisingly identical to those obtained with the continuous outcome.

The de jure analyses now exhibit a moderate divergence between CCMV [1.63 (0.18) $p=0.006$], ACMV [1.58 (0.18) $p=0.011$], and NCMV [1.46 (0.20) $p=0.061$]. Particularly, the treatment effect estimated under NCMV is not statistically significant. Among de facto analyses, the treatment effect estimated under N-LOCF [1.52 (0.19) $p=0.027$] is still greater than that estimated under N-BOCF [1.39 (0.21) $p=0.117$]. The difference is caused by 13 additional clinical responders for Test against 3 for PCB. As with the continuous outcome, N-DO5 [1.48 (0.21) $p=0.059$] and N-AE5 [1.47 (0.19) $p=0.047$] provide similar treatment-effect estimates although only N-AE5 is statistically significant.

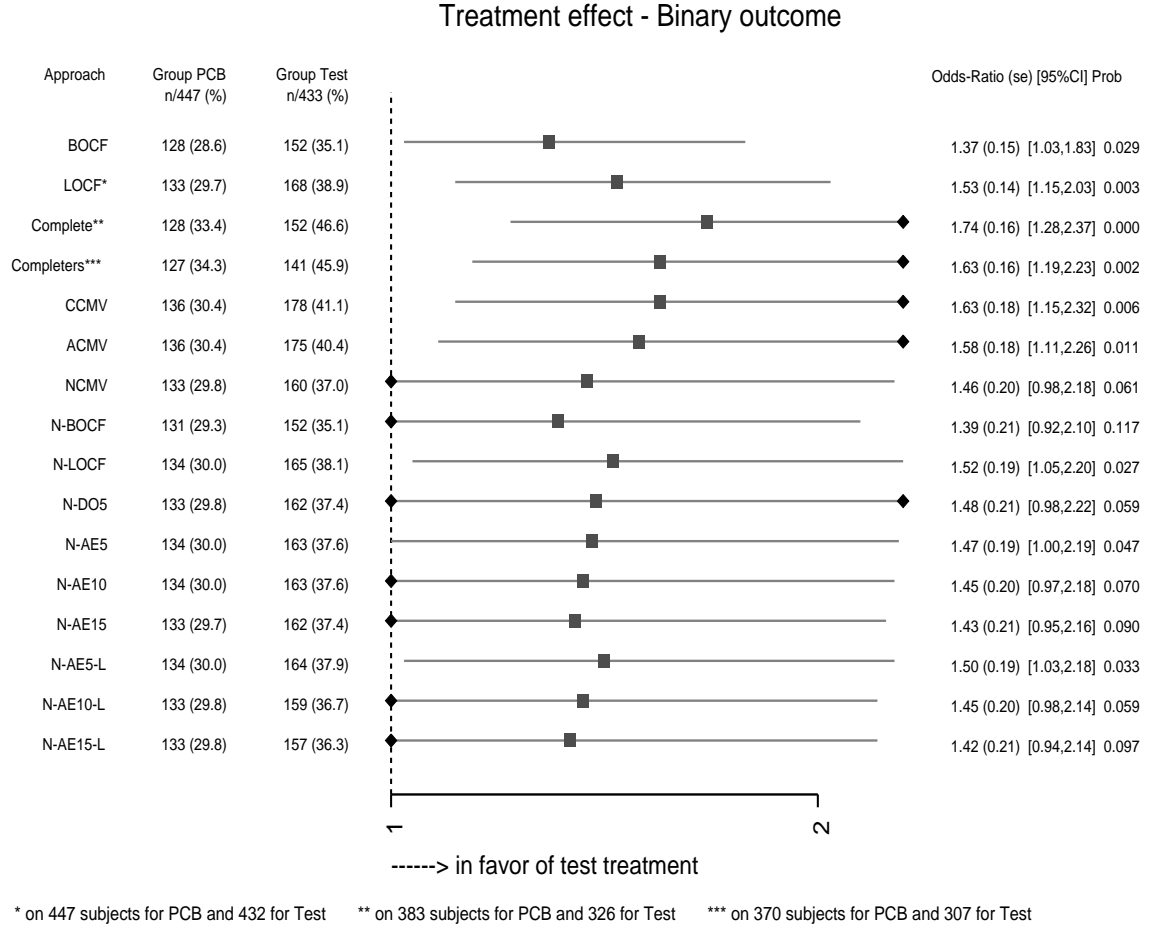


FIGURE 3 – Treatment-effect estimates obtained with the binary outcome at visit 4.

Whereas inferences obtained so far with the binary outcome remain coherent with the continuous case, the N-AEx strategies exhibit divergences. The penalty increase only implies a slight decrease of odds-ratios as observed under N-AE10 [1.45 (0.20) $p=0.070$] and N-AE15 [1.43 (0.21) $p=0.090$] relative to N-AE5, although the p -values change to statistically non-significant. The slight impact on odds-ratios is caused by a floor effect on the numbers of clinical responders. From N-AE5 to N-AE15, these numbers vary from 134 to 133 for PCB and from 163 to 162 for Test. A similar impact of penalty increase is observed in N-AE5-L [1.50 (0.19) $p=0.033$], N-AE10-L [1.45 (0.20) $p=0.059$], and N-AE15-L [1.42 (0.21) $p=0.097$]. Another consequence of the floor effect is that the treatment effect under the N-AEx and the N-AEx-L strategies remains in favor of Test (i.e., odds-ratio >1) whatever the Δ value. Otherwise, the limit values for statistical non-significance are $\Delta = -6$ under N-AEx and $\Delta = -8$ under N-AEx-L.

5.3 Sensitivity analysis

Table 4 compares the results obtained from the corrected marginal approach which are laid out in Figure 2 and Figure 3 with those based on full analyses of the pattern-conditional and the marginal effects.

TABLE 4 – Results obtained from the corrected marginal approach and based on full analyses of the pattern-conditional and the marginal effects.

Outcome	Drop-out strategies	pattern-conditional effect	Corrected marginal approach	Marginal effect
CONTINUOUS Difference (s.e.) p	CCMV	5.03 (1.50) 0.001	5.03 (1.63) 0.002	5.15 (1.63) 0.002
	ACMV	4.78 (1.49) 0.001	4.78 (1.63) 0.004	4.94 (1.63) 0.003
	NCMV	4.85 (1.51) 0.001	4.85 (1.67) 0.004	5.05 (1.67) 0.003
	N-BOCF	1.52 (1.79) 0.397	1.52 (1.89) 0.424	1.83 (1.89) 0.272
	N-LOCF	3.92 (1.59) 0.014	3.92 (1.72) 0.024	4.04 (1.72) 0.020
	N-DO5	3.21 (1.67) 0.055	3.21 (1.78) 0.073	3.31 (1.78) 0.064
	N-AE5	3.13 (1.63) 0.055	3.13 (1.74) 0.073	3.30 (1.74) 0.059
	N-AE10	2.21 (1.71) 0.198	2.21 (1.80) 0.222	2.38 (1.80) 0.187
	N-AE15	1.23 (1.83) 0.503	1.23 (1.90) 0.520	1.42 (1.90) 0.458
	N-AE5-L	3.58 (1.58) 0.023	3.58 (1.73) 0.040	3.85 (1.73) 0.027
	N-AE10-L	2.82 (1.62) 0.082	2.82 (1.76) 0.111	3.06 (1.76) 0.084
	N-AE15-L	2.04 (1.68) 0.223	2.04 (1.80) 0.259	2.25 (1.80) 0.213
BINARY Odds-Ratio (s.e.) p	CCMV	1.63 (0.15) 0.001	1.63 (0.18) 0.006	1.62 (0.18) 0.007
	ACMV	1.58 (0.15) 0.002	1.58 (0.18) 0.011	1.57 (0.18) 0.013
	NCMV	1.46 (0.15) 0.010	1.46 (0.20) 0.061	1.42 (0.20) 0.084
	N-BOCF	1.39 (0.15) 0.028	1.39 (0.21) 0.117	1.26 (0.21) 0.272
	N-LOCF	1.52 (0.15) 0.005	1.52 (0.19) 0.027	1.47 (0.19) 0.039
	N-DO5	1.48 (0.15) 0.008	1.48 (0.21) 0.059	1.42 (0.21) 0.089
	N-AE5	1.47 (0.15) 0.009	1.47 (0.19) 0.047	1.42 (0.19) 0.071
	N-AE10	1.45 (0.15) 0.013	1.45 (0.20) 0.070	1.37 (0.20) 0.114
	N-AE15	1.43 (0.15) 0.017	1.43 (0.21) 0.090	1.34 (0.21) 0.159
	N-AE5-L	1.50 (0.15) 0.007	1.50 (0.19) 0.033	1.51 (0.19) 0.031
	N-AE10-L	1.45 (0.15) 0.012	1.45 (0.20) 0.059	1.44 (0.20) 0.064
	N-AE15-L	1.42 (0.15) 0.020	1.42 (0.21) 0.097	1.39 (0.21) 0.115

For the continuous outcome, an overall look at results across strategies indicates that the point estimates based on the marginal effect are always greater than those given by the pattern-conditional effect. This is a consequence of the imbalance in the treatment assignment across patterns. The over-estimation by the marginal effect is moderate with a variation from 0.10 under N-DO5 to 0.31 under N-BOCF. The marginal effect also produces greater standard errors because of the integration of the distribution of patterns. Consequently, the corrected marginal approach, which combines terms given by the

pattern-conditional and the marginal effects, yields the most conservative inferences. Of note, the inferences remain coherent across approaches since none of the p -values switches from significant to non-significant, and vice-et-versa, from one approach to the other.

The coherence across approaches observed on the continuous outcome does not hold on the binary scale because of the non-linear metric used to characterize treatment-effect magnitude. Except in N-AE5-L, the point estimates of the marginal effect are now lower than those of the conditional-pattern effect. Of the twelve strategies, height exhibit discrepancies in inferences between the pattern-conditional and the marginal effects. However, the inferences obtained from the corrected marginal approach and marginal effect analysis are almost coherent since only N-AE5 exhibits a discrepancy with 1.47 (0.19) $p=0.047$ for the first and 1.42 (0.19) $p=0.071$ for the second.

5.4 Subgroup analysis

Table 5 provides the treatment-effect estimates in the subgroups of subjects who dropped out for AE (DOAE) and other reasons than AE (DOnotAE).

In the subgroup DOAE, only NCMV on the continuous scale yields a statistically significant effect of Test. The other drop-out strategies produce treatment effects that are either clinically moderate or counterbalanced by the penalties for drop-out in the N-AEx and N-AEx-L strategies. The subgroup DOnotAE exhibits opposite results since all the drop-out strategies, except N-BOCF on the continuous scale, yield statistically significant effects of Test. In this subgroup, the continuous outcome means by treatment group are all positive for Test and negative for PCB, whereas the rates of clinical responders are always greater in the Test group.

These results show no evidence of treatment effect in the subjects who dropped out for AE whereas the subjects who dropped out for other reasons unambiguously benefited from Test. The lack of treatment effect in subjects suffering from AE indicates that subjects actually suffered from inconveniences influencing there VAS score which were not relievable by a treatment indicated to the targeted disease (i.e., fibromyalgia). This also underlines the relevance of the N-AEx-L strategies in our case study.

Further attention is needed to interpret the results after BOCF imputation. They exhibit non-null continuous outcome means by treatment group and some clinical responders although BOCF assigns a 0 value and a non-clinical response. In fact, these outcome values come from the subjects of pattern 4 who dropped out after visit 3 and had their drop-out

TABLE 5 – Treatment-effect estimates in the subgroups of subjects who dropped out for AE and other reasons than AE († indicates a treatment effect in favor of PCB).

Subgroups	Drop-out strategies	Mean		Diff (se)	p	Freq (%)		OR (se)	p
		PCB	Test			PCB	Test		
DOAE	BOCF	-.10	2.2	2.3 (1.4)	0.098	0/44 (0.0)	4/96 (4.2)	5.23 (1.37)	0.228
	LOCF	5.0	7.6	2.6 (2.9)	0.372	5/44 (11.4)	18/96 (18.7)	1.9 (0.56)	0.257
	CCMV	9.1	13.2	4.0 (2.7)	0.130	7/44 (15.9)	26/96 (27.1)	1.74 (0.47)	0.234
	ACMV	7.9	11.4	3.5 (2.5)	0.172	7/44 (15.9)	24/96 (25.0)	1.57 (0.47)	0.335
	NCMV	.45	8.5	8.1 (2.1)	<.001	5/44 (11.4)	10/96 (10.4)	0.85 (0.56)	0.768 †
	N-BOCF	-1.4	-2.4	-1.0 (2.8)	0.722	3/44 (6.8)	4/96 (4.2)	0.59 (0.72)	0.460 †
	N-LOCF	3.4	6.2	2.8 (2.7)	0.308	6/44 (13.6)	15/96 (15.6)	1.10 (0.51)	0.854
	N-DO5	-1.9	.36	2.3 (2.8)	0.418	5/44 (11.4)	12/96 (12.5)	1.02 (0.55)	0.966
	N-AE5	-.64	.62	1.3 (2.9)	0.667	5/44 (11.4)	12/96 (12.5)	1.02 (0.55)	0.966
	N-AE10	-4.9	-5.3	-.38 (3.2)	0.907 †	5/44 (11.4)	11/96 (11.5)	0.93 (0.55)	0.903 †
	N-AE15	-9.6	-12.0	-2.1 (3.7)	0.576 †	4/44 (9.1)	11/96 (11.5)	1.19 (0.59)	0.770
	N-AE5-L	5.7	5.2	-.47 (2.1)	0.822 †	5/44 (11.4)	13/96 (13.5)	1.07 (0.54)	0.902
	N-AE10-L	1.8	.07	-1.7 (2.2)	0.431 †	4/44 (9.1)	8/96 (8.3)	0.83 (0.61)	0.755 †
	N-AE15-L	-2.2	-5.1	-2.9 (2.3)	0.206 †	4/44 (9.1)	6/96 (6.3)	0.63 (0.63)	0.467 †
DOnotAE	BOCF	-1.0	6.7	7.7 (2.7)	0.006	1/33 (3.0)	7/30 (23.3)	9.0 (0.95)	0.021
	LOCF	-5.7	12.3	18.1 (3.6)	<.001	1/33 (3.0)	9/30 (30.0)	17.7 (1.02)	0.005
	CCMV	-2.2	15.8	18.0 (2.6)	<.001	2/33 (6.0)	11/30 (36.7)	8.71 (0.78)	0.006
	ACMV	-3.3	14.8	18.1 (3.3)	<.001	2/33 (6.0)	11/30 (36.7)	8.71 (0.78)	0.006
	NCMV	-5.8	12.0	17.8 (2.3)	<.001	1/33 (3.0)	9/30 (30.0)	12.7 (0.96)	0.008
	N-BOCF	-.89	.54	1.43 (4.4)	0.747	1/33 (3.0)	7/30 (23.3)	7.10 (0.92)	0.034
	N-LOCF	-5.1	11.1	16.2 (3.5)	<.001	1/33 (3.0)	9/30 (30.0)	12.7 (0.96)	0.008
	N-DO5	-9.5	7.25	16.7 (3.7)	<.001	1/33 (3.0)	9/30 (30.0)	12.7 (0.96)	0.008
	N-AEx	-3.3	14.3	17.7 (3.5)	<.001	2/33 (6.0)	10/30 (33.3)	7.62 (0.79)	0.010
	N-AEx-L	-2.8	14.6	17.4 (3.5)	<.001	2/33 (6.0)	10/30 (33.3)	7.62 (0.79)	0.010

visit repositioned to visit 4 for analysis.

The results obtained under NCMV in the subgroup DOAE also deserve explanations since the effect of Test is statistically significant with the continuous outcome [8.1 (2.1) $p<.001$] whereas analysis on the binary scale yields a slight advantage in favor of PCB [OR=0.85 (0.56) $p=0.768$]. Relative to CCMV, the means by treatment group under NCMV decrease by -8.6 for PCB and -4.7 for Test. Such impact was expected from Figure 1, which exhibits a marked divergence in the mean profiles between the drop-out patterns 1–4 used in NCMV and pattern 5 of completers used in CCMV. In parallel, the numbers of clinical responders decrease by $-2/44$ (-4.5%) for PCB and $-15/96$ (-15.6%) for Test. These contradictory results between the continuous and the binary outcomes are only caused by a scale effect. The substantial gain for PCB on the continuous scale corresponds to few

additional clinical responders because the threshold of 30% improvement is not reached. These results in the subgroup DOAE explain the divergence between the inferences under NCMV based on the continuous and the binary outcomes in the whole sample.

Under the N-AEx and N-AEx-L strategies, the influence of the penalty for drop-out in the subgroup DOAE is unsurprisingly quasi-linear on the continuous scale. Conversely, the impact on the binary outcome is weak because of the floor effect. In the subgroup DNotAE, the penalty has no impact at all since subjects are imputed assuming MAR.

6 Discussion

Planning a PMM-MI analysis requires one to question and address all the aspects of implementation. In our case study, MMRMs with full interaction of treatment by visit are used for imputation. This modeling strategy has driven the definition of patterns, which are based on drop-out at each visit. In this context, the large number of parameters required and the sharing of parameters often emerge as a limiting factor as the pattern sizes do not allow to accommodate all the potentially useful information. Although the same pattern is supposed to represent the same kind of subjects, it is likely that the correlation structure in subjects who dropped out for AEs differs from that of subjects who dropped out for other reasons. Furthermore, while we estimate parameters separately for each pattern, we also assumed, in de jure analyses, that the covariance is the same across treatment groups.

Another important aspect of our implementation concerns the derivation of the conditional pattern distributions in (8). The method is based on J2R which is implemented by default in our software solution. After drop-out, patient's mean response distribution is now that of the PCB group. Such a change may be seen as a worst-case scenario in terms of reducing any treatment effect since drop-out subjects on Test will lose the effect of their period on treatment. With additional programming, J2R can be replaced by another method such as *copy increments in reference* [17]. This alternative performs imputation so that the subject's mean profile tracks that of the mean profile in the PCB group, but starting from the benefit already obtained with Test if a given subject was on Test before drop-out.

Planning a PMM-MI analysis also requires one to cautiously examine characteristics of patterns like the pattern sizes (e.g., majority completer versus spread over many patterns) and the proximity between patterns (e.g., are there major jumps in mean and/or variance structure from one pattern to the neighboring one). Imputation based on the pattern

of completers may contribute to undue overestimation of the treatment effect and may impose inappropriately low uncertainty. In our case study, imputation is based on the drop-out patterns only. These patterns have homogeneous sizes and their mean profiles exhibit stable trends over time. The consequence is that, outside the effect of the free distribution of values present under NFD, drop-out strategies have a moderate impact on treatment-effect estimates, while a reasonable amount of uncertainty is allowed. It is also important to note that, under NFD, early drop-out has less impact, compared to MAR, than late drop-out assuming commonly used covariance structures. It is a concept that several statisticians may find counter-intuitive when reflecting upon de facto analysis.

In our NFD implementation, the present corresponds to the first unobserved visit and the distribution of present values are defined relative to the drop-out visits in the closest neighboring patterns. It is important to assess the impact of these choices, especially as we cannot rule out that efficacy assessments at these visits combine other aspects related to tolerance. In our case study, we considered that including the drop-out visits in the analysis does not bias efficacy evaluation and, on the contrary, provides valuable information for treatment-effect estimation. This statement is based on the comparisons of analysis results with and without the drop-out visits, wherein the moderate difference observed was in line with what was observed in the mean profiles of patterns.

7 Concluding remarks

As already stated and exemplified in [17], PMM-MI methods permit the implementation of various assumptions on missingness mechanisms. Among them, NFD offers an appealing setting to formalize beliefs on drop-out and address a large spectrum of scientific issues. Strategies can easily be tailored according to plausible clinical scenarios. We show that NFD offers powerful solutions to tackle major drawbacks of well-known single-imputation concepts such as BOCF and LOCF. We also show that strategies may aim to balance the benefits against risks in the assessment of the treatment effect. Of note, in our case study it was very feasible to refine the penalties for drop-out by incorporating some information on the type of adverse event leading to drop-out as given in Section 2.2. So, the penalties are adjusted according to the impact of AE on the subject. An important aspect of our NFD-PMM-MI approach is that it aims more at analysing an outcome whose value is corrected in some circumstances related to drop-out, rather than directly challenging the statistical properties of a primary analysis. As such, the approach fits more into the concept

of estimand than the field of sensitivity analysis.

The possibility to draw inferences on the marginal treatment effect is an important aspect of our implementation. This characteristic is compatible with analyses of confirmatory RCTs. In this setting, we point to the issue of imbalance in the treatment allocations across patterns very clearly and we suggest a pragmatic way out in the presence of moderate imbalance. These cases of imbalance are common in PCB-controlled trials in chronic pain when the test treatment causes early drop-out due to adverse reactions. Another possibility allowed by MI-based methods is the investigation of the impact of drop-out strategies in subgroups of interest using the individual profiles averaged over imputation.

Any candidate method to analyse confirmatory RCTs implies the possibility to plan all the aspects of analysis before breaking the blind. In this manuscript, we highlight several points to consider that can be addressed during analysis preparation. We also show how to implement the approach using a freely available existing program. Accordingly, there is now no reason not to consider an NFD assumption and PMM-MI methods for primary and sensitivity analyses in confirmatory RCTs.

References

- [1] Molenberghs G, Kenward MG. *Missing Data in Clinical Studies*, Wiley & Sons, 2007.
- [2] Little RJA. Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* 1993 ; **88** : 125-134
- [3] National Academy of Sciences. *Prevention and Treatment of Missing Data in Clinical Studies*. National Academies Press ; 2010.
- [4] Bunouf P, Molenberghs G, Grouin J-M, Thijs H. A SAS program combining R functionalities to implement pattern-mixture models. *J. STAT. SOFTW.* 2015 ; **68**.
- [5] Bunouf P, Grouin J-M, Molenberghs G, Koch G. Analysing intensive longitudinal data after summarization at landmarks : an application to daily pain evaluation in a clinical trial. *J. R. Statist. Soc. A* 2012 ; **175**, Part 2 : 513-534.
- [6] Bunouf P, Grouin J-M, Molenberghs G. Analysis of an incomplete binary outcome derived from frequently recorded longitudinal continuous data : application to daily pain evaluation. *Statist. Med.* 2012 ; **31** : 1554-1571.
- [7] Carpenter JR, Kenward MG. Missing data in clinical trials – a practical guide. *Bir-*

mingham : *National Health Service Co-ordinating Centre for Research Methodology* ; 2008.

- [8] Molenberghs G, and Michiels B, Kenward MG, Diggle PJ. Missing data mechanisms and pattern-mixture Models. *Statistica Neerlandica* 1998 ; **52** : 153-161.
- [9] Kenward M, Molenberghs G, and Thijs H. Pattern-mixture models with proper time dependence. *Biometrika* 2003 ; **90** : 53-71.
- [10] Molenberghs G, Beunckens C, Sotito C, Kenward MG. Every missing not at random model has got a missing at random counterpart with equal fit. *Journal of the Royal Statistical Society, Series B* 2008 ; **70** : 371–388.
- [11] Rubin DB. *Multiple Imputation for nonresponse in Surveys*, Wiley & Sons ; 1987.
- [12] Thijs H, Molenberghs G, Michiels B, Verbeke G, Curran D. Strategies to fit pattern-Mixture Models. *Biostatistics* 2002 ; **3** : 245-265.
- [13] Hogan JW, Roy J, Korkontzelou C. Tutorial in biostatistics. Handling drop-out in longitudinal studies. *Statistics in medicine* 2004 ; **23** : 1455-1497.
- [14] Yan X, Lee S, Li N. Missing data handling methods in medical device clinical trials. *Journal of Biopharmaceutical Statistics* 2009 ; **19** :1085-1098.
- [15] Ratitch R, O’Kelly M, Tosiello R. Missing data in clinical trials : from clinical assumptions to statistical analysis using pattern mixture models. *Pharmaceutical Statistics* 2013 ; **12** : 337-347.
- [16] Gallo P, Chuang-Stein C. A note on missing data in non-inferiority trials. *Drug Information Journal* 2009 ; **43** : 469-474.
- [17] Carpenter JR, Roger J, and Kenward, MG. Analysis of longitudinal trials with missing data : A framework for relevant, accessible assumptions, and inference via multiple imputation. *Journal of Biopharmaceutical Statistics* 2013 ; **23** : 1352-1371.

Appendix : Software implementation

This appendix provides the reader with additional information to implement the corrected marginal approach using an existing SAS program which combines R functionalities [4]. This program produces pooled treatment-effect estimates under NCMV, ACMV, NCMV, and NFMV using the three-stage Rubin’s method. Minimal knowledges in SAS and R are needed to implement the additional functionalities.

In the existing program, estimation of parameters and pooled analysis are performed in SAS whereas MI is entirely carried out under R. The program structure of the R script is a first limitation as MI is conducted under the same identifying restrictions in all subjects. The reason is that MI specifications, including the identifying restriction, are entered once in the SAS program and are then exported to R. To allow for individual imputation schemes, the loop for subject imputation must be transferred from the R script into the SAS program. Of note, this causes much more calls of R from SAS when running the program.

Next, the existing program does not impute missing covariates and first outcome values. This functionality must be inserted in the R script using a syntax similar to that used for imputation of the other missing values. Another modification in the R script concerns the numbering of patterns and the number of observed visits per pattern which match by default. In our case study, subjects with complete outcome profile may pertain to patterns 4 or pattern 5 although four visits are observed in subjects of both patterns. Information on how to decouple these two parameters is available in Section 7.4 in [4].

To end, the syntax to perform the direct pooled analysis in the SAS program must be replaced by a syntax to perform pooled analyses per pattern and then average treatment effects accross patterns. Using the notation in Section 4.2.3, we suggest the following algorithm :

1. Use SAS Proc MIXED or SAS Proc LOGISTIC to fit outcome values at visit 4 per pattern by imputation and estimate \mathbf{d} and standard errors ;
2. Use SAS Proc FREQ to estimate the pattern probabilities in $\mathbf{\Pi}$;
3. Use SAS Proc MIANALYZE to estimate \mathbf{d}^* and W^* which contain pooled treatment effects and within-imputation standard errors per pattern ;
4. Use SAS Proc CORR to estimate the between-imputation covariance matrix B^* ;
5. Use R or SAS to calculate the pattern-averaged parameters W^\dagger , B^\dagger , V^\dagger and produce the results.